

**caBIG Architecture Workspace**  
**Organization and Scope of Work**  
**April 16, 2004**

**Overview**

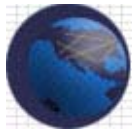
The Architecture workspace believes the focus of caBIG should be to create a data-sharing infrastructure, or "Data Grid", as distinguished from a processor-sharing infrastructure, or "Computation Grid." These respective types of capabilities are not mutually exclusive, and computational grid features can be added if and when they are deemed a high priority.

To address the many architectural issues surrounding the design and construction of a data grid, the Arch. Workspace will begin by partnering with the liaisons from other Workspaces to develop specific use cases. The use cases that will be needed can be grouped into a series of sub-topic areas, and the Arch. Workspace has organized a corresponding set of sub-groups to deal with these areas.

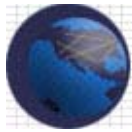
This document describes the scope of work to be covered in each subgroup. In each area, use case development and reference implementation activities will be conducted in conjunction with projects going on in other Workspaces.

**Subgroups**

1. Interface Architecture. Use cases and requirements for exposed APIs, query interfaces, message interfaces, data/metadata structures, and data grid services.
  - a. Capture API and messaging interface use cases and requirements, recommend development standards that meet the requirements.
  - b. Describe formats and structures for data and accompanying metadata delivery from a caBIG-compatible system.
  - c. Liaison with other workspaces:
    - i. Vocab/CDE liaisons: Scott Oster (Ohio State); Robert Robbins (Hutch)
    - ii. Clin. Trials. Mgmt. liaisons: Frank Manion (Fox Chase); Derek Walker (Hutch)
    - iii. Tissue and Pathology liaisons: Rakesh Nagarajan (Wash U) ; Arnie Miles (Georgetown)
    - iv. Integrated Cancer Research liaisons: Simon Lin (Duke); Joel Saltz (Ohio State); Steve Moore (Georgetown);
2. Information Architecture. Conceptual model formalisms, metadata requirements, data mapping and relationship formalisms, ID mgmt, data and model change control.
  - a. Establish standards for information object model definition and representation i.e. UML.
  - b. Metadata requirements and representation
  - c. Terminology mapping to NCI Thesaurus and other standards



- d. Universal identifier specification and management
- e. Change management, versioning and history maintenance
- f. Liaison with other workspaces
  - i. Vocab/CDE liaisons: Scott Oster (Ohio State); Robert Robbins (Hutch);
  - ii. Clin. Trials. Mgmt. liaisons: Tahsin Kurc (Ohio State);
  - iii. Tissue and Pathology liaisons: Rakesh Nagarajan (Wash U)
  - iv. Integrated Cancer Research liaisons: Frank Manion (Fox Chase) ; Joel Saltz (Ohio State);
- 3. System Architecture. Runtime technologies, service advertisement, execution of grid queries, message brokering, data grid workflow management and execution.
  - a. Data Grid framework and toolkit exploration
    - i. Globus
    - ii. OGSA-DAI
    - iii. Shibboleth
    - iv. caCORE
    - v. Other...
  - b. Message brokers
  - c. Workflow
- 4. Security and Access control. Authentication and authorization strategy; patient privacy protection; honest broker services. This functionality needs to be part of the runtime environment, but the group felt that it should be addressed as a distinct set of requirements and may need special attention.
  - a. Security requirements, use cases for protected limiting access
  - b. De-identification technology
  - c. HIPAA compliance
  - d. Liaison with other workspaces
    - i. Clin. Trials. Mgmt. liaisons: Jim Harrison (Pittsburgh); Tahsin Kurc (Ohio State);
    - ii. Tissue and Pathology liaisons: Steve Moore (Georgetown); Joel Saltz (Ohio State); Jim Harrison (Pittsburgh); Tahsin Kurc (Ohio State);
    - iii. Integrated Cancer Research liaisons: Steve Moore (Georgetown); Joel Saltz (Ohio State);
    - iv. Data Sharing and Intellectual Capital Liaisons: Robert Robbins (Hutch);
- 5. Software Development: Best practices and Standards. Identify and promote common development standards for building caBIG compatible software. The goals of the group are to explore current software development practices underway at national cancer



centers with the goal of documenting a set of minimal requirements for building caBIG compatible software.

- a. Best practices:
    - i. Use Case development
    - ii. Modeling
    - iii. User interface design
    - iv. Prototyping and iterative development
    - v. Software Quality Assurance, including unit, functional and system testing
    - vi. Documentation
  - b. Training materials/tutorial development. Developed in conjunction with Training workgroup. Architecture will play consultative and role, and provide primary content where needed.
  - c. Implementation support and assistance for Domain Workspaces
  - d. Liaison with other workspaces
    - i. Training liaisons: Jim Harrison (Pittsburgh);
6. Reference Implementations. At periodic intervals, a reference implementation that instantiates all the components should be constructed. These reference implementations should be drawn from pilot projects in the domain workspaces.
7. Extended Groups

## **Communications**

### Within Architecture

There will be separate “sub-groups” folders created on the forum site relating to each of the Architecture Workspace sub-groups. Therefore each area will have a thread for each sub-group. Dynamic Project web pages are coming up soon. There will be regularly scheduled conference calls that will be scheduled much in advance so that the various members will be able to plan ahead to attend the calls. We are also getting closer to being able to use Centra for our Webinar/Webcast sessions.

### Cross-Cutting with other Workspaces

Designated inter-workspace liaisons will participate in inter-workspace tele-conferences and meetings. Each subgroup must send at least one liaison to these events. Liaisons will report back to the Architecture Workspace as a whole.